



# Применение методов искусственного интеллекта для оценивания качества речи при проведении речевой реабилитации

Костюченко Евгений Юрьевич

Начальник научно-инженерного центра «Интеллектуальные системы доверенного взаимодействия» ЦК НТИ «Технологии доверенного взаимодействия» Томского государственного университета систем управления и радиоэлектроники, Кандидат технических наук, Доцент

# Качество речи

Качество речи - величина, характеризующая оценку звучания речи с точки зрения применения для решаемых задач.

Характеризуется тремя основными аспектами:

Разборчивость речи - относительное количество (в процентах) правильно принятых элементов (фонем, слогов, слов, фраз) артикуляционных таблиц.

Узнаваемость речи (голоса диктора) - величина, характеризующая степень сохранения индивидуальных признаков голоса диктора после внесения изменений и оцениваемая как верно классифицированных по принадлежности диктору речевых фрагментов.

Естественность речи - величина, характеризующая оценку соответствия звучания оцениваемой речи естественному произношению. Количественно определяется как доля верных классификаций при проведении попарных сравнений естественной и оцениваемой речи.

# Актуальность работы

Злокачественные новообразования головы и шеи являются сложной медико-социальной проблемой современной онкологии, причиной заболеваемости и смертности трудоспособного (до 60 лет) населения. Хирургическое вмешательство в условиях комбинированного лечения полости рта и ротоглотки неизбежно приводит к выраженной деформации периферического речевого аппарата. Возникают дефекты звукопроизношения и нарушения ритма акустического восприятия речи у пациента. Это приводит к инвалидности пациента.

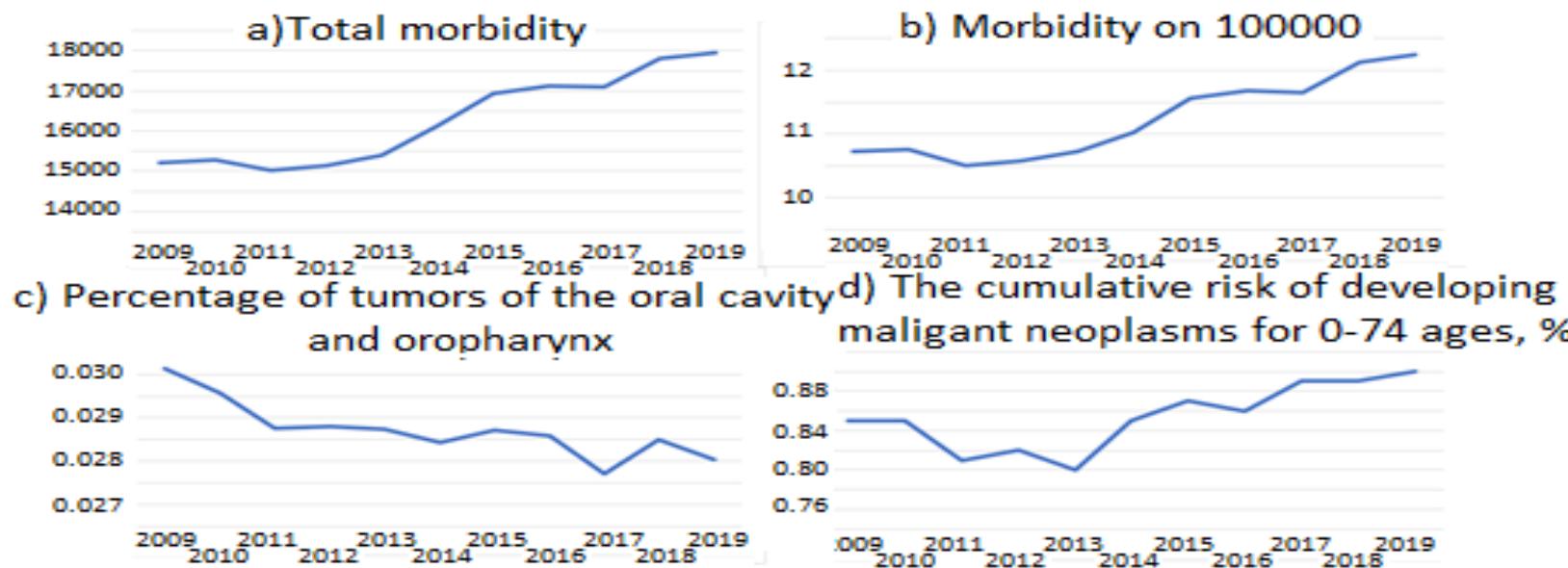


Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия

# Актуальность работы

Статистика диагностики и лечения опухолей речеобразующего тракта, полости рта и рса указывает на увеличение количества больных. Оперативное вмешательство в процессе ухудшает качество жизни пациента. Методы оценивания качества речи, применяемые при реабилитации в основном были субъективными.

Создание методологии оценивания качества речи позволяет перейти к автоматизированным объективным количественным оценкам, дает возможность пациенту проводить часть тренировок самостоятельно, сокращает сроки голосовой и речевой реабилитации.



Злокачественные новообразования в России в 2019 г. (заболеваемость и смертность) Под ред. А.Д. Красинского, А.О. Шахзаде. МНИОИ им. П.А. Герцена – «НМИЦ радиологии» Минздрава России. – М.: МИАМ, 2020. – 128 с.

# Актуальность работы

Оцениванием качества речи занимаются несколько ключевых центров:

№	Центр	Лидеры	Направление
1.	University of Southern California	Shrikanth Narayanan	Исследование различных видов речевых пациентов
2.	Chung Shan Medical University	Wen-Shin Hsu,Guang-Tao Lin	Критерии разборчивости, узнаваемости и естественности при дизартрии
3.	Massachusetts Institute of Technology, Cambridge	Geoffrey S.Meltzner	Критерии разборчивости и естественности у электроларинкса
4.	Institute of Health and Society, Speech and Language Sciences, Newcastle University, Newcastle upon Tyne, UK	Nick Miller	Различные критерии оценивания разборчивости
5.	Yamagata University	Kazuhiro Kondo	Оценивание разборчивости речи с использованием
6.	Cognitive Systems Lab, University of Bremen, Bremen, Germany	Christian Herff, Lorenz Diener, Miguel Angrick, Tanja Schultz	Генерация естественной, узнаваемой речи на основе мозговой активности
7.	University of Wisconsin, Madison	Raymond D. Kent	Отдельные критерии разборчивости речи в речевых расстройствах
8.	Санкт-Петербургский Федеральный исследовательский центр Российской академии наук	Карпов А.А., Ронжин А.Л.	Отдельные критерии оценивания эмоциональной выразительности, имеющие выход на оценивание естественности
9.	Санкт-Петербургский государственный университет	Ляксо Е.Е.	Применение отдельных критериев качества речи в нарушениях у детей

# Актуальность работы

Проанализированные работы учитывают только один критерий качества для анализа здорового диктора. Например, вместо оценивания узнаваемости решается прямая идентификации/аутентификации, вместо оценивания общей естественности распознавания отдельные особенности эмоциональной окраски речи. Таким образом, отсутствует метод оценивания качества речи при проведении речевой реабилитации.



Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия

# Актуальность работы

Отдельно можно выделить аспекты связи направления исследования и информационной безопасности.

Разборчивость — используется с точки зрения оценивания защищенности помещений, обработке речевой информации, их аттестации на соответствие требованиям и т. д.

Узнаваемость речи (голоса диктора) имеет ключевое значение для систем биометрической аутентификации и идентификации личности.

Естественность речи выступает важным индикатором при выявлении фальсификаций и гла之城 подделок голоса.



Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия



# Цель

**Демонстрация методологии оценивания качества при проведении речевой реабилитации для повышения эффективности и универсальности по сравнению с использованием отдельных методов за счет комплексного объективного подхода.**



Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия



# Объект и предмет исследования

Объект исследования — речевой сигнал в норме и при нал изменений (связанных с патологией, оперативным вмешательством, попыткой модификации и т.д.).

Предмет исследования — получение объективных количественных оценок качества речевого сигнала по выбранным критериям качества на базе методов параметризации речевого сигнала, анализа и интерпретации полученных параметров.



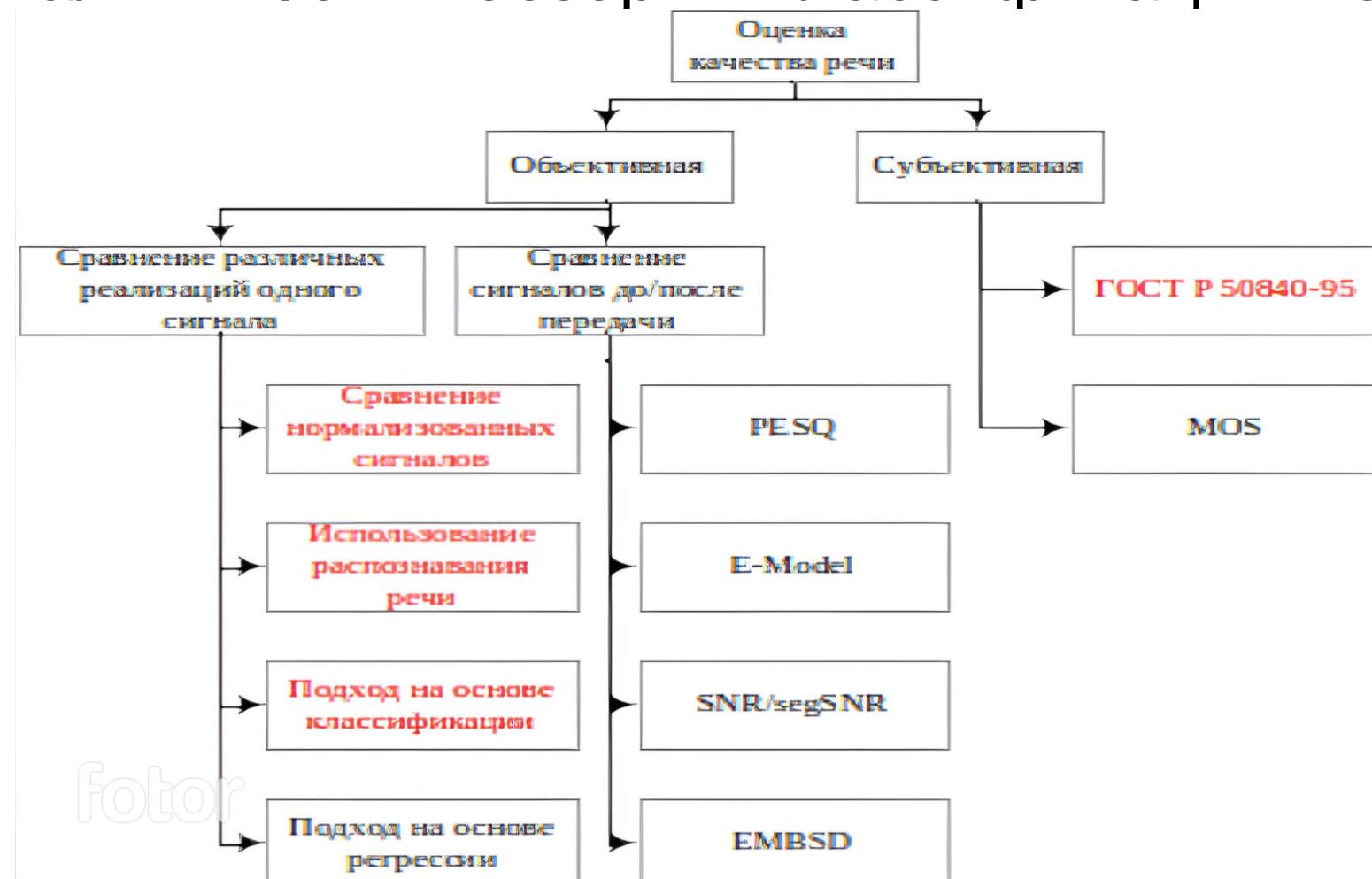
## Аналитический обзор

Проведен поиск по ключевой фразе Speech Quality Estimation. Найдено 59 работ, индексируемых Web of Science (ключевые слова), 868 работ, индексируемых Scopus (ключевые слова), 207 работ, индексируемых Google Scholar (слова в названии). При этом публикации равномерно распределены по годам последнего десятилетия([63, 49, 40, 41, 42, 41, 30, 36, 49, 46] статей), что подтверждает актуальность данного исследования.





# Аналитический обзор — классификация методов



Полученная классификация служит основой для формирования отдельных элементов методологии оценивания качества речи

## Аналитический обзор

Проведена классификация объективных и субъективных методов оценивания качества речи. Объективные позволяют получать стабильные не зависящие от аудитора оценки, вызывающие затруднения при интерпретации. Субъективные понятны, просты в получении, но этот процесс требует участия субъектов-аудиторов, причем получаемые оценки полностью зависят от них и, как правило, изменяются при замене команды аудиторов. Кроме того, сбор такой команды в практике не возможен.



## Аналитический обзор

Объективные методы оценивания разделены на 2 категории. Одна из категорий – это методы сравнения сигналов в рамках телекоммуникационных систем. Это приводит к сравнению, по-сути, одного сигнала с двумя его копиями, внесенными в процесс передачи помехами. Вторая категория направлена на сравнения различных реализаций сигнала для различных вариантов произнесения одной фразы. Второй подход вариативен – сигналы принципиально отличаются по длительности, однако имеют более широкое применение за счет отсутствия ограничений на сравнение одного сигнала на различных стадиях обработки.

# Методология

**Методология — совокупность методов и способов реализации, представлены в виде последовательных этапов их применения для достижения основной цели.**

**Методология определяет компоненты научного исследования – объект, задачи исследования, а также представляет четкую последовательность действий исследователя при решении задач.**

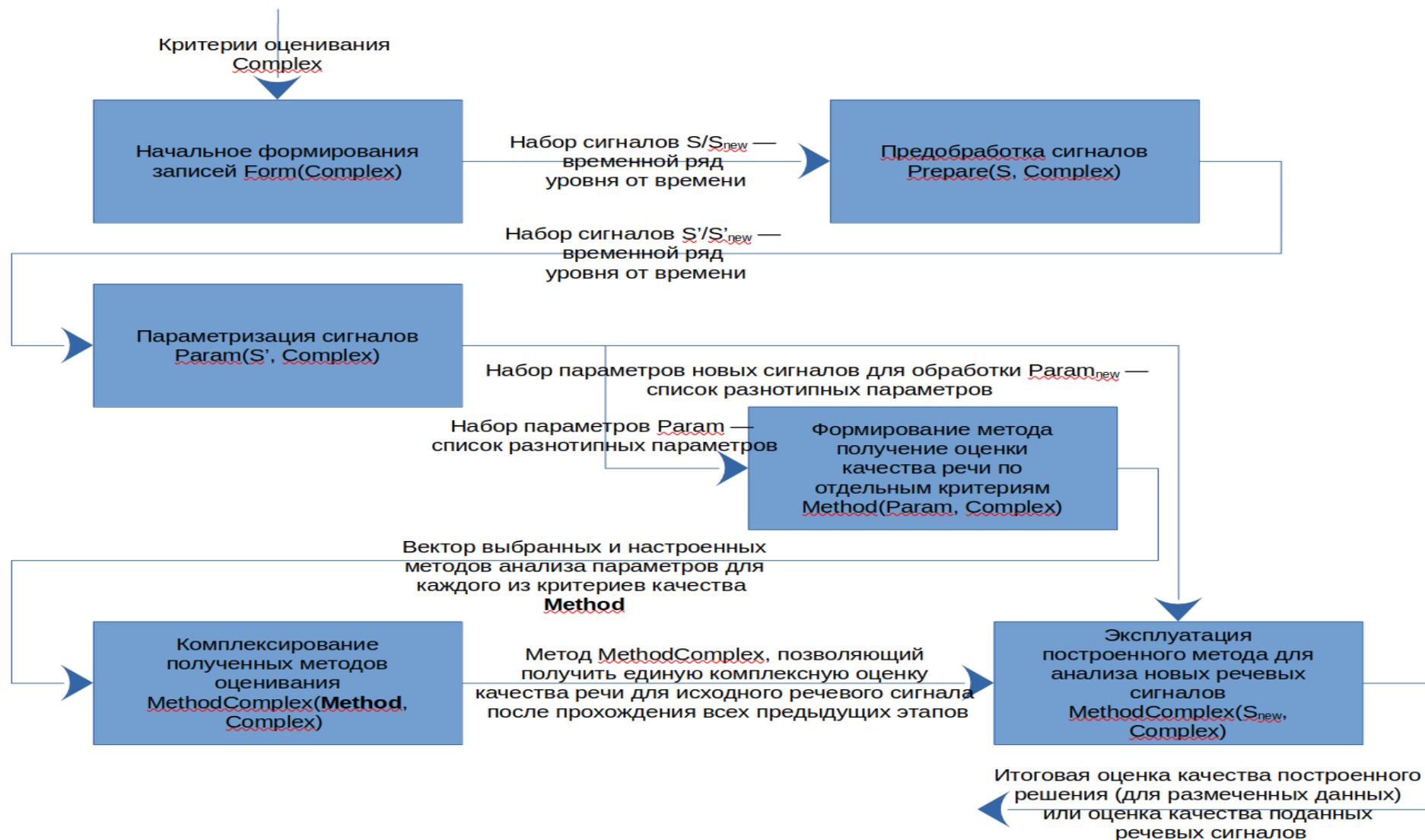
**Выделяют наиболее важные точки приложения методологии.**

- 1. Постановка задачи.**
- 2. Построение предмета исследования.**
- 3. Построение научной теории.**
- 4. Проверка полученного результата с точки зрения его истинности.**





# Методология





# Выбор критериев оценивания (0)

Для формированием набора данных необходимо заранее определить те из критериев разборчивости, естественности и узнаваемости, по которым будет производиться оценивание. Выбор задается через задание целого трехбитового параметра Complex=Complex<sub>d</sub>=inr<sub>b</sub>, отвечающего за задание множества комплексируемых параметров оценивания. Задание может осуществляться в бинарном или десятичном виде, в десятичном Complex<sub>d</sub> ∈ [1;7]. По умолчанию принимается оценивание с точки зрения использования всех критериев оценивания Complex<sub>d</sub>=7.



# Формирование набора (1)

$S$  — множество сигналов. Данное множество сигналов разбивается на 3 взаимодополняющих подмножеств:

$S_{i+}$ ,  $S_{i-}$ ,  $S_{ij}$  — эталонные наборы, отражающие эталонную норму, максимально некачественные сигналы и сигналы в промежуточном состоянии с точки зрения оценивания разборчивости (intelligibility).

$S_{n+}$ ,  $S_{n-}$ ,  $S_{nj}$  — эталонные наборы, отражающие эталонную норму, максимально некачественные сигналы и сигналы в промежуточном состоянии с точки зрения оценивания естественности (naturalness).

$S_{r+}$ ,  $S_{r-}$ ,  $S_{rj}$  — эталонные наборы, отражающие эталонную норму, максимально некачественные сигналы и сигналы в промежуточном состоянии с точки зрения оценивания узнаваемости (recognition).

Где  $j$  – номер сеанса реабилитации



# Формирование набора (1)

В рамках оценивания предполагается выбор или формирование метрик качества Q речи, обладающих свойством монотонного изменения от состояния после операции к состоянию до операции для штатного проведения реабилитации:

Если  $Q_{complex+} > Q_{complex-}$ , то  $Q_{complex+} > \dots > Q_{complex2} > Q_{complex1} > Q_{complex-}$

Если  $Q_{complex-} > Q_{complex+}$ , то  $Q_{complex-} \rightarrow Q_{complex1} > Q_{complex2} > \dots > Q_{complex+}$



# Предобработка (2)

Включает в себя следующие подэтапы:

1. Шумоочистка записей Denoise(S). Применяется для повышения качества записей, может использован при необходимости и применен как ко всему набору, так и к любым отдельным фрагментам вплоть до отдельных записей.
2. Нормирование сигнала по уровню громкости LevelNorm(S) — направлен на приведение обрабатываемых записей к одинаковым условиям, направлен на устранение вариативности параметров уровня микрофона при записи, расстояния от него и т.п.
3. Расширение набора записей Extend(S, Complex). Расширение набора при недостаточном объеме или при наличии несбалансированности по отдельным критериям. Использоваться как методы аугментации, так и генерации новых записей на основе имеющихся. При этом, полученные файлы дополняют исходный набор, и, в зависимости от способа генерации и использованных при получении записей, также распределяются по поднаборам  $\{S_{i+}, S_{i-}, S_{ij}, S_{n+}, S_{n-}, S_{nj}, S_{r+}, S_{r-}, S_{rj}\}$ .



## Шумоочистка или выравнивание по шуму

1. Спектральное вычитание. Анализ спектра сигнала и шума: на этапе паузы или речи оцениваются коэффициенты, отражающие спектральную структуру шума. На основе этого определяется шум, который затем вычитается из общего спектра сигнала. Прост в реализации, но может привести к появлению артефактов, известных как «музыкальный шум», из-за неточной оценки шума.
2. Вейвлет-преобразование предлагает многочастотный анализ сигнала. Сигнал раскладывается на составляющие в разных частотных полосах и временных интервалах, после чего коэффициенты, соответствующие шуму, пороговой обработкой обнуляются или ослабляются. Эффективен для нестационарных шумов, так как позволяет локально во времени и по частоте управлять компонентами сигнала.
3. Адаптивная фильтрация использует дополнительный опорный сигнал шума. Адаптивный фильтр динамически подстраивает свои коэффициенты, чтобы вычесть из основного сигнала оценку текущего уровня шума. Этот метод высокоэффективен в ситуациях, когда шум коррелирован и есть возможность его записать отдельно.
4. Статистические методы, например, фильтр Винера - проектируются для минимизации среднеквадратической ошибки между очищенным и целевым сигналами, основываясь на статистических характеристиках и речи, и шума. Обеспечивает хорошее качество, когда эти характеристики известны и могут быть достаточно точно оценены.
5. Методы на основе машинного обучения, в частности, глубоких нейронных сетей. Модели обучаются на больших наборах данных восстанавливать из зашумленных речевые сигналы их чистые версии. Такие системы способны эффективно подавлять разнородные типы шумов, включая нестационарные, что превосходит традиционные алгоритмы, особенно при работе с сильными помехами.



## Нормирование по громкости

Нормирование громкости подразумевает изменение уровня громкости всего аудиофайла или отдельного сегмента, основываясь на фиксированных расчетных показателях. Этот процесс осуществляется на основе таких базовых критериев, как:

- Пиковый уровень: Сигнал масштабируется таким образом, чтобы его самый высокий амплитудный пик стал равным определенному уровню (например, 0 dBFS). Этот критерий гарантирует, что аудиофайл не будет превышать допустимые уровни громкости.
- Средний уровень громкости (RMS - Root Mean Square): Расчет сигналов проводится с учетом средней мощности звука, что позволяет учитывать более общий динамический диапазон. Нормирование способствует тому, чтобы восприятие громкости файла стало однородным и комфортным.

Однако этот подход не всегда помогает в устраниении резких различий громкости между разными частями записи. Например, если в аудиофайле есть перепады громкости в песнях или фрагментах, однозначное нормирование не исправляет внутренние динамические изменения.





## Выравнивание по громкости

Выравнивание громкости — это более комплексный процесс, включающий динамику обработки сигнала. Главная цель — обеспечение одинакового уровня громкости в течение времени воспроизведения. Метод может включать:

- Изменение диапазона с учетом восприятия: Автоматическое управление уровнями громкости в тихих и громких частях с учетом равномерности восприятия пользователем.
- Динамическая компрессия: Низкие уровни сигнала усиливаются, а высокие корректируются, обеспечивая более сбалансированное восприятие. Например: использование алгоритма автоматического управления динамическим диапазоном (Dynamic Range Control) или корректора для предотвращения "разрывов громкости" между соседними элементами воспроизведения.
- Профилирование по слуховым нормам: Современные технологии могут учить особенности человеческого слуха, что позволяет определить, в каких частотах изменения громкости наиболее заметны, например алгоритмы громкости на основе модели восприятия (Perceptual Loudness Models).

В отличие от нормирования, выравнивание громкости применяется именно для уравнивания громкости в разных частях аудиофайла: например, таких как голосовой трек в фильме или музыка.



Аугментация Кипяткова И. С., Кагиров И. А. Аналитический обзор методов решения проблемы малых наборов данных при создании систем автоматического распознавания речи для малоресурсных языков //Информатика и автоматизация. – 2022. – Т. 21. – №. 4. – С. 678-709.

Таблица 1. Результаты применения методов аугментации обучающих данных для обучения систем автоматического распознавания речи

Методы аугментации	Изменяемые параметры	Работы	Относительное сокращение ошибки распознавания слов (WER, %)
Речевые данные	высота голоса/ темп речи/ громкость речи	[25]	2,00
	добавление случайных значений к речевым признакам	[26]	0,30
	изменение спектрограммы	[27]	27-46 (в зависимости от используемых данных и применения модели языка)
	преобразование голоса (voice conversion)	[46]	31,36
	синтез речи	[47]	60,40
Текстовые данные	использование текстовых данных другой предметной области	[39]	2,70 – 3,90 (в зависимости от объема данных)



# Метод переноса знаний (применение предобученных сетей, transfer learning)

Метод обучения ИНС, когда знания нейросети, которая была обучена на задаче, переносятся на другую задачу. Этот метод используется в том случае, если обучающих данных для целевой задачи мало, но имеется большой обучающий корпус для другой смежной задачи.

Таблица 2. Результаты применения метода переноса знаний для малоресурсных языков

Работа	Нецелевой язык (языки)	Целевой язык	Методы	Относительное сокращение ошибки распознавания слов (WER, %)
[54]	нидерландский	африкаанс	Гетероскедастический линейный дискриминантный анализ и адаптация к речи диктора	34,34 <sup>‡</sup>
[55]	английский + путунхуа	амхарский	Перенос параметров предобученной модели для инициализации параметров целевой модели	37,73
[56]	английский	сенека	Перенос знаний	25,05
			Перенос знаний + аугментация	41,13

Продолжение Таблицы

Работа	Нецелевой язык (языки)	Целевой язык	Методы	Относительное сокращение ошибки распознавания слов (WER, %)
[57]	английский	швейцарский немецкий	Перенос параметров предобученной модели без заморозки параметров	-2,70
			С заморозкой параметров 1 и 2 слоя ИНС	9,46
[58]	амхарский, тигриня, оромо, воламо	Языки из GlobalPhone + речевые данные близкого языка	Переносились веса скрытых слов, а веса выходного слоя дообучались для каждого целевого языка на данных этого языка	2,61
				-0,48
				0,87
				1,81
[59]	уйгурский	Языки из GlobalPhone	Перенос параметров многоязычной модели в целевую модель	33,21
[60]	ибан	Языка из TIMIT	Многозадачное обучение	1,90-5,90 (в зависимости от количества обучающих данных)
[66]	зулу, коса, сесото, тswana	M-BERT, 104 языка	5,53	
			2,09	
			50,07	
			41,00	

# Параметризация (3)

Выделяемые параметры могут быть новыми параметрами, имеющими семантическое значение выделения, или другой формой представления речевого сигнала для применения специфических обработки, не свойственных в явном виде речевому сигналу.

1. В качестве множества первой группы параметров и вариантов их получения могут выступить параметров openSMILE ComParE\_2016, openXBOW BoAW, DeepSpectrum DenseNet и А выделяемые с помощью специализированных библиотек, таких как OpenSmile и аналогичные.

Категория параметров	Конкретные примеры
Низкоуровневые дескрипторы	Энергия сигнала, громкость
Спектральные параметры	Mel-/Bark-/Octave-спектры, MFCC, PLP-CC, форманты
Параметры качества голоса	Джиттер (Jitter), шиммер (Shimmer)
Просодические и тональные параметры	Частота основного тона (Pitch), классы высоты тона (CENS)
Параметры на основе КЛП	Линейное предсказание (LPC), Линейный спектральный анализ (LSP)
Дескрипторы формы сигнала	Автокорреляция, кепстральный анализ

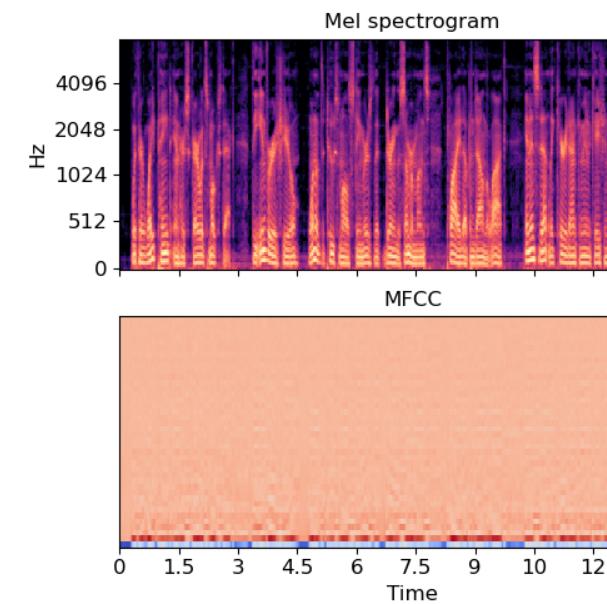
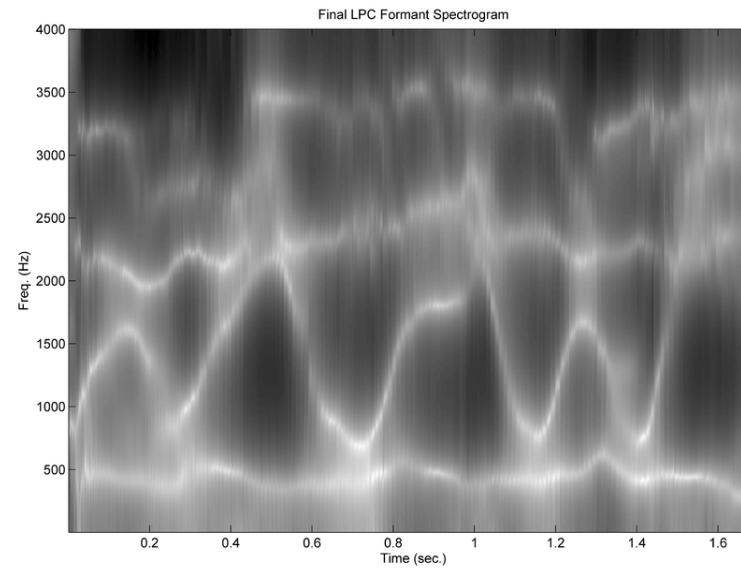
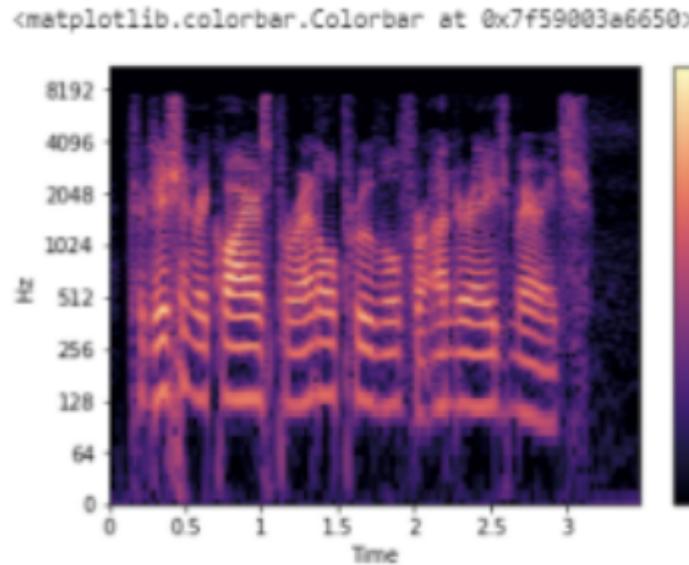
1. Amiriparian S., Sokolov A., Aslan I., et al. On the Impact of Word Error Rate on Acoustic-Linguistic Speech Emotion Recognition: An Update for the Deep Learning Era. arXiv: 2104.10121



# Параметризация (3)

2. Под второй группой параметров подразумевается подготовка сигнала к виду, пригодному для использования методов обработки изображений, в частности, с применением предобученных сверточных нейронных сетей, для анализа, например, представления исходного сигнала в спектрально-временной форме. Для применения таких методов необходимо провести предварительное преобразование исходного сигнала для представления его в виде образа для возможности применения соответствующих методов. Это также может быть представлено как вариант параметризации.

Используются Mel-/Bark-/Octave-спектры, MFCC, PLP-CC, другие разновидности спектров.



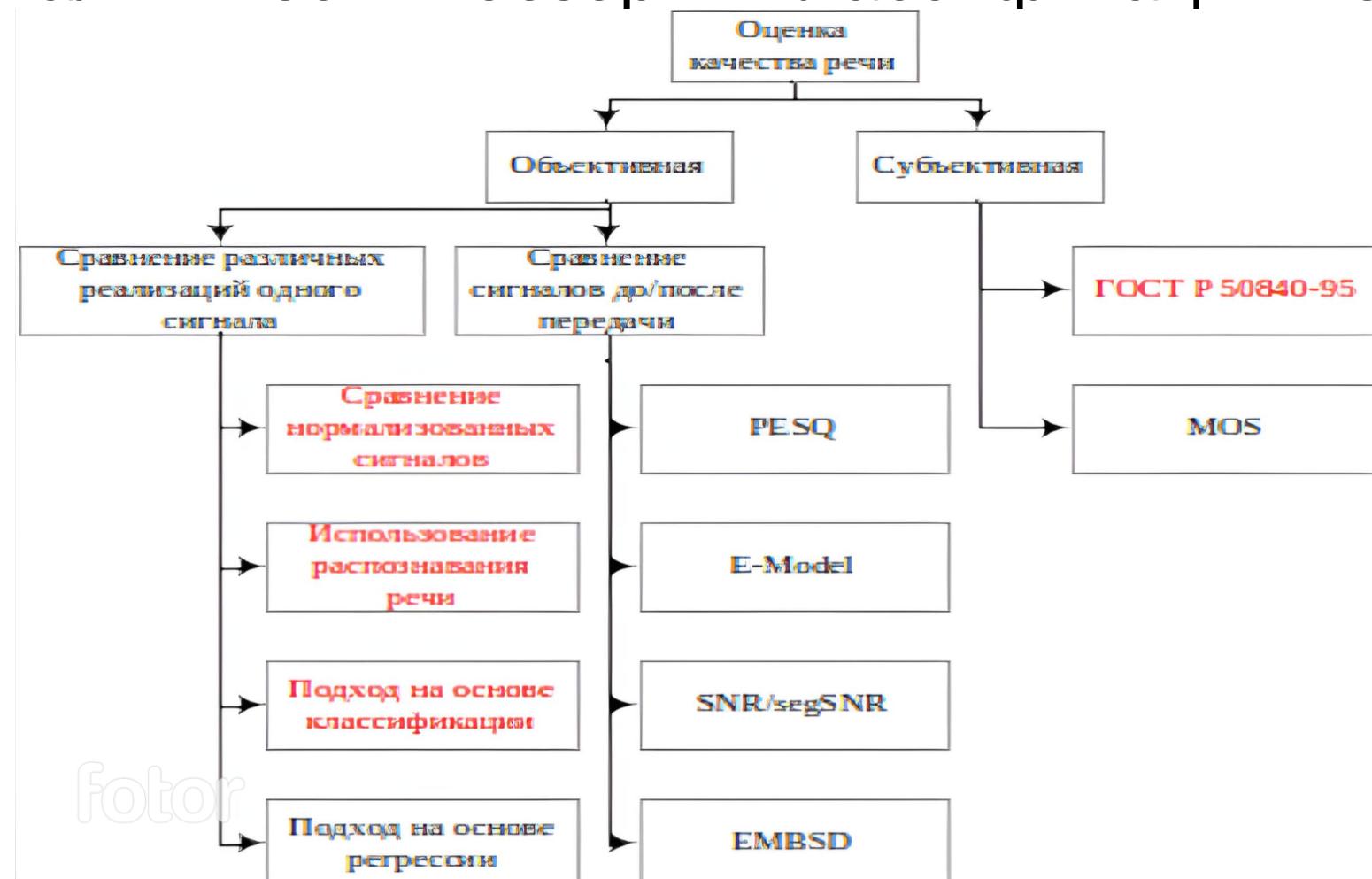


# Формирование метода оценивания качества речи по отдельным критериям (4) - Группы методов

Методы оценивания	Описание
1. Субъективные экспертные.	В явном виде в данной работе не рассматриваются, более того, работа направлена на их замену методами, исключающими субъективизм и участие экспертов, однако, группа оставлена для существующих методов (со всеми их недостатками) в разрабатываемую методологию и их описание для сравнения в качестве аналога в тех областях решения практических задач, где автоматизация ранее не использовалась.
2. Основанные на адаптации существующих методов оценивания качества речи при ее передаче по каналам связи.	Например, перцептивная оценка качества речи (PESQ) (метрика качества определяется исходя из методом).
3. Основанные на применении определения различных расстояний между классами.	Используется расстояние между экземплярами эталонного и негативного наборов (метрика определяется на основе расстояний между экземплярами внутри/между эталонным и негативным наборами).
4. Основанные на применении различных методов анализа данных применительно к решению задачи классификации исходного набора на эталонный и негативный классы.	Метрика качества — мера уверенности принадлежности эталонному или негативному классу.
5. Основанные на анализе речи, специфичные для отдельных критериев качества речи, а также отдельных форм представления речевого сигнала.	Применение методов распознавания речи для оценивания разборчивости речи (метрика качества определяется на основе доли верно распознанных речевых единиц). Полученный результат для оценивания передается методам группы 3.
6. Использование регрессионных методов.	Оценка — результат применения методов регрессии к размеченным векторам входных параметров.



# Аналитический обзор — классификация методов



Полученная классификация служит основой для формирования отдельных элементов методологии оценивания качества речи

## ГОСТ Р 50840-95

5 Аудиторов. Слоговая разборчивость (1000 таблиц по 50 слогов). Фразовая разборчивость (100 таблиц по 50 фраз)  
Качество речи по селективным признакам. Узнаваемость чистота записи.

Таблица Б.999

ТЕЛЬ	ТАЦ	ЗЮЛ	ЖЕНЬ	ЧТЕН	СУН	ЧЕНЬ	ЦЕН	СЫС	ЛАП
б			б	д		о	з	з	б
ЛУП	ШАЛ	НЕРЬ	ДИП	НИТ	БЁМ	ЧЁЦ	НЯМЬ	ЗЫСЬ	ШАПЬ
з	б	в	з	з	д	сч	з	з	з
ЦУС	МЯПЬ	ПЮФЬ	ФИС	ПЕС	СЁТ	ЩАЦ	ХАСЬ	ЧТАТ	СТЭТ
тезнь	РЫК	ЗЫМ	ХАФ	НАСЬ	МИМ	ДАСЬ	ДЯСЬ	МАМ	ГАК
РЕР	ФАНЬ	БЫХ	ФСЕТ	ДОШ	ЛЮСЬ	НОК	ДОС	ФОЙ	ЛИФЬ

Таблица Б.1000

ПЯНЬ	ТРАМ	ФСЯШ	РЯЛ	ЕЙ	СИЛЬ	КАФЬ	МЕЛЬ	ЖАС	ЗДЕЙ
в	я	б	г	в	сч	жь	и д	з	с
НЕФ	ЧАЦ	НОПЬ	НАК	БУФЬ	ЩИК	ЗИШ	ШИТ	РИРЬ	ПАФ
о в					д	б		д	
ЖЁФЬ	РАМ	БЕН	ТРЁК	ДЕМ	СВЁТ	ЗНОП	ЗЕЧ	ТРУНЬ	ЗУТЬ
в									
ПРЕФ	СВОН	ЧТАЙ	ДАН	МУЛЬ	РЫН	ДЕЗНЬ	ШТАК	КРОХ	КИЦ
в з	з	б				сч	ь	и д	
ФСАС	НИСЬ	ДЛИП	ТАНЬ	НОЦ	ХАЗНЬ	ЩЕЛ	САЩ	ВЁТ	ТРЕТ

Таблица Д.1

Номер п/п	Фразы	Номер п/п
1	В бухте с моря налетел ветерок	26
2	Дно у реки хорошее	27
3	Мальчик побежал к лагерю	28
4	Сигнал тревоги поднял отряд	29
5	Прошло всего несколько минут	30
6	Штурман просил продолжать разворот	31
7	Ледяная вода сводила руки	32
8	Лошадь тихонько пофыркивала	33
9	Часовой стоял на берегу	34
10	Яркая луна освещала лыди	35
11	В доме стало холодно	36
12	В печке дотлевали последние угли	37
13	Солнце медленно опускалось за море	38
14	Лошадь неслась вскачь	39
15	Дети остались сиротами	40
16	Снегсыпал в лицо	41
17	Маяк был уже близко	42
18	Ветер стучал в ставни	43
19	На столе мигала лампа	44
20	Лодка замедлила ход	45
21	Она стала тихой и молчаливой	46
22	Около дома навалены сугробы	47
23	Громко залаяли собаки	48
24	Песня лилась плавно и звучивно	49
25	Командир корабля одобрил решение	50

Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия



## Методы при оценивании в рамках канала связи

Метод	Тип	Принцип	Ключевое применение
<b>MOS</b>	Субъективный	Оценка живыми слушателями по 5-балльной шкале	Эталонная проверка качества
<b>PESQ</b>	Объективный (полноссыл.)	Сравнение исходного и искаженного сигнала с учетом психоакустики	Тестирование сетей 3G, VoIP
<b>POLQA</b>	Объективный (полноссыл.)	Улучшенное сравнение для широкополосного и полноспектрального звука	Тестирование сетей 4G/5G (VoLTE/VoNR)
<b>ViSQOL</b>	Объективный (полноссыл.)	Моделирование слуховой системы через сравнение нейросвиповых спектрограмм	Широкополосные кодеки (Оригинальные видеоконференции)
<b>E-Model</b>	Параметрический	Расчет R-фактора на основе параметров канала	Планирование и мониторинг с реальном времени
<b>ITU-T P.563</b>	Объективный (бесссыл.)	Анализ только принятого сигнала без эталона	Мониторинг качества в работающей сети

- + Формальные, стандартизованные. На выходе объективные количественные оценки.
- Используются для сравнения идентичных сигналов. Требуют выравнивания и не имеют стохастический характер. Завязаны на канал связи.



## Нормирование по времени — простое передискретизирующее выравнивание

Проводится линейное преобразование осцилограммы для выравнивания продолжитель сравниваемых единиц через передискретизацию сигнала. Использует меха интерполяции.

- + Очень прост и быстр.
- ± Терпимо для коротких единиц — фонем, слогов.
- неприменим для продолжительных высказываний.

Применен в рамках [Kostyuchenko E. et al. Evaluation of the speech quality during rehab after surgical treatment of the cancer of oral cavity and oropharynx based on a comparison fourier spectra //International Conference on Speech and Computer. – Cham : Springer Intern Publishing, 2016. – С. 287-295.] для последующего сравнения через расстояния.



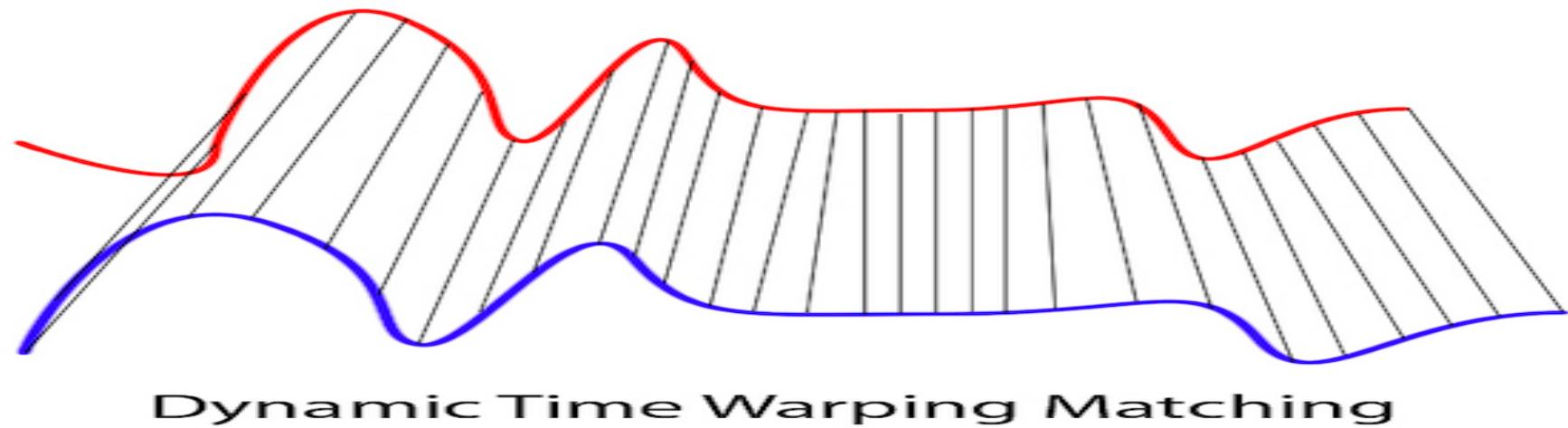
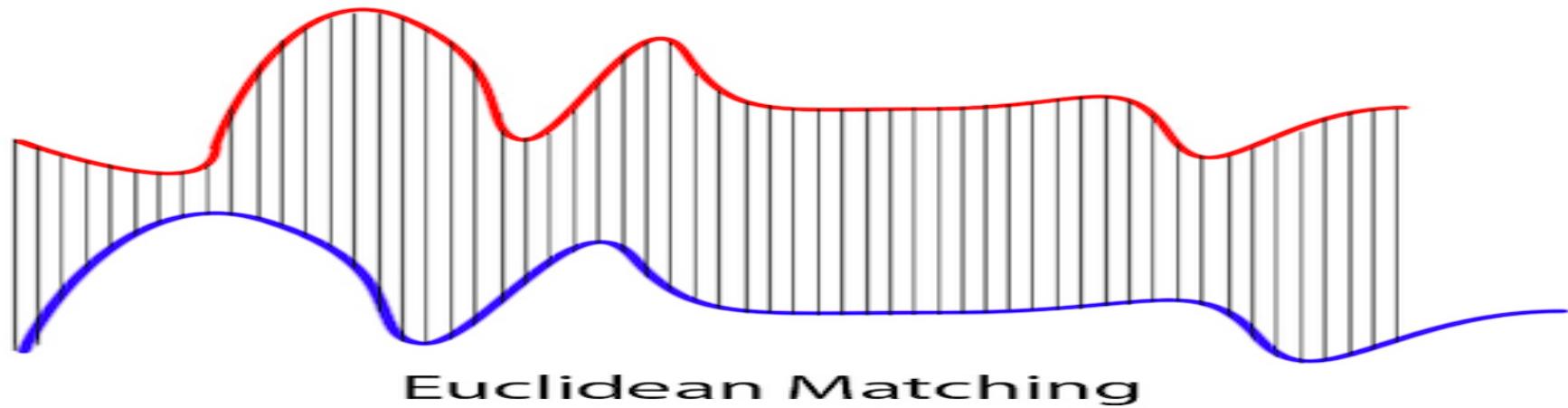
# Нормирование по времени — корреляционное выравнивание с учетом количества фонем.

- Проводится пофонемная сегментация эталонного слога с учетом известного количества фонем и фонетического состава.

Алгоритм / Подход	Краткое описание	Сильные стороны и примечания
<b>Алгоритм на основе измерения спектральных переходов (STM)</b>	Вычисляет меру спектральных переходов для обнаружения границ между фонемами. Реализован в том числе для работы в реальном времени на мобильных устройствах.	Прямо предназначен для точного выделения границ между фонем. Проверен в сравнении с другими методами сегментации и разметкой экспертов. Хороший результат в исследовательских задачах.
<b>Современные инструменты сегментации (WhisperX, NeMo)</b>	Мощные современные инструменты, использующие нейросетевые модели (трансформеры, конформеры). Изначально предназначены для задач диаризации и транскрипции.	Высокая скорость и точность. Адаптированы или использованы для различных задач, включая извлечение детализированных описаний фонетических гранул (дифоны).

- Проводится подбор положения точек границ в оцениваемом сигнале так, чтобы корреляция линейного выравнивания сегментов по отношению к эталонному была максимальной. Kostyuk et al. Correlation normalization of syllables and comparative evaluation of pronunciation quality in speech rehabilitation //International Conference on Speech and Computer. – Cham : Springer International Publishing, 2017. – С. 262-271. Можно максимизировать коэффициент для любого представления (осцилограмма, огибающая мощности, спектrogramма). Позволил достигнуть среднего уровня корреляции более 0,8 для осцилограмм сигналов.

Нормирование по времени — применение DTW (динамическая трансформа временной шкалы).

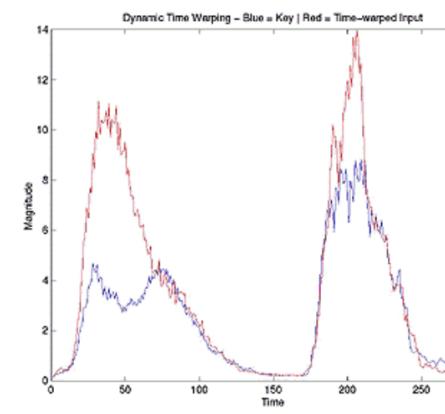
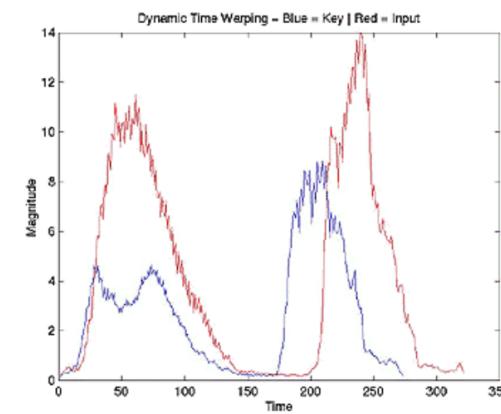
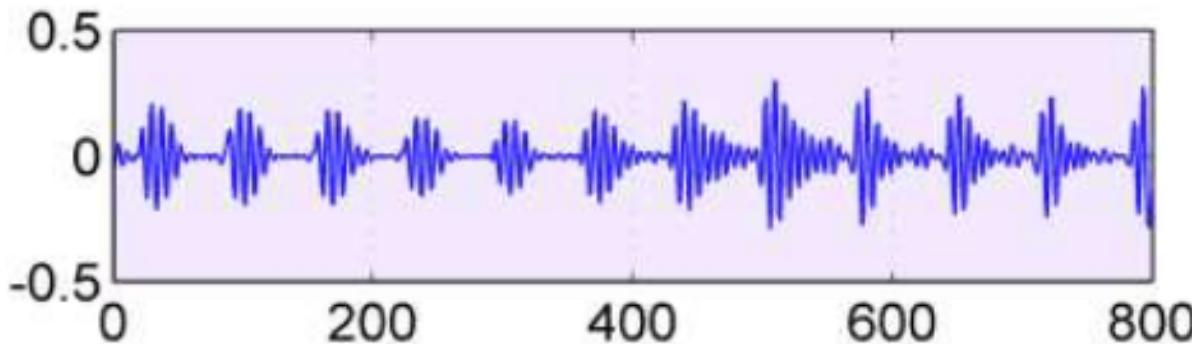


## Нормирование по времени — применение DTW (динамическая трансформации временной шкалы).

DTW решает следующую задачу: как сравнить две временные последовательности, которые имеют разную переменную скорость или быть смещены во времени? DTW находит оптимальное выравнивание между двумя последовательностями, "растягивая" и "сжимая" их вдоль времени, чтобы суммарное расстояние между сопоставленными точками было минимальным. Алгоритм динамического программирования — имеет квадратичную сложность.

Плохо применим для осцилограмм из-за случайного характера количества волн в единицах (нестабильности ЧОТ) даже на чисто вокализованных участках.

Требует использования к огибающим выравненных по относительной мощности участков.  
Использование ФНЧ.



## Сравнение различных реализаций одного сигнала

1. Применение одного из стандартов.
2. Использование расстояний. Сравнение расстояний между эталонных сигналов и парой эталонный-оцениваемый сигнал.

Могут использоваться разные варианты расстояний — манхэттенова, Евклида, его обобщения - махalanобиса, Минковского. В [Kostyuchenko al. Evaluation of the speech quality during rehabilitation after surgical treatment of the cancer of oral cavity and oropharynx based on a comparison of the spectra //International Conference on Speech and Computer. – Cham : Springer International Publishing, 2016. – С. 287-295.] использован подбор метрики Минковского для минимизации расстояния между спектрами сигналов. Позволило сформировать одну из первых количественных мер объективного оценивания качества.



## Применение методов классификации

Строится классификатор в зависимости от решаемой задачи, например:

1. Для разборчивости — непроблемные фонемы, список проблемных фонем по отдельности известен. Мера качества — усредненная метрика уверенности принадлежности корректному классу/доля верных ответов [Pekarskikh S., Kostyuchenko E., Balashov A. Evaluation of Speech Quality Through Recognition and Classification of Phonemes //Symmetry. – 2021. – №. 12. – С. 1447.].
2. Узнаваемость — 2 класса ответов — пары сигналов принадлежат одному диктору или нет. Ориентирован на «сиамский подход». Мера качества — усредненная мера уверенности в правильных ответах/доля верных ответов [Kostyuchenko E., Rakhmanenko I., Lapina M. Evaluation of a method for measuring speech quality based on an authentication approach using a correlation criterion //2021 International Conference on Intelligent Environments (IE). – IEEE, 2021. – С. 1-7.].
3. Естественность — 2 класса ответов — пары сигналов принадлежат одной группе или разным. Мера качества — усредненная мера уверенности в правильных ответах/доля верных ответов [Litvinova N., Laptev P., Kostyuchenko E. Application of Machine Learning Techniques for Generated Speech Identification //2024 32nd Telecommunications Forum (TELFOR). – IEEE, 2024. – С. 1-4.].



TUSUR

University

## Модели распознавания речевых сигналов

[<https://alphacepheli.com/nsh/2025/04/18/russian-models.html>]

Dataset	Vosk 0.54	Vosk 0.54 LODR	Nemo RNNT Fastconformer	Nemo Parakeet TDT V3	Nemo Canary V2	Whisper Large V3 Transformers	Whisper v3 Turbo	Whisper Podlodka Turbo	GigaAM2 RNNT	GigaAM2 CTC + LM	Vosk Small Streaming 0.54	T-one CTC + LM Streaming	Vikhr Borealis
Аудиокниги АЦ	<b>1.2</b>	1.3	8.2	6.9	12.0	5.8	6.5	6.9	4.4	3.4	4.1	5.8	8.4
Ru Librispeech	9.4	9.0	11.2	10.6	19.8	9.5	9.7	9.3	5.2	<b>4.4</b>	14.4	6.2	5.9
CommonVoice 12.0	6.1	5.6	5.9	5.3	8.7	5.5	6.2	5.2	<b>2.6</b>	2.9	11.2	5.5	2.9
Golos Crowd	3.1	3.0	2.7	3.9	9.2	14.7	14.5	11.1	2.5	<b>2.2</b>	5.5	5.6	8.0
Golos Farfield	6.2	5.9	7.1	7.6	15.6	17.6	18.7	10.9	4.4	<b>4.1</b>	10.1	12.5	11.3
Sova устройства	11.6	11.4	7	16.2	19.8	15.9	16	14.0	<b>5.6</b>	8.3	14.7	10.1	14.5
Телевещание	16.6	16.2	22.6	18.0	21.3	17.9	18.2	19.7	14.4	<b>13.8</b>	19.8	19.5	22.7
Медицина	15.6	15.4	19.2	13.4	16.9	13.8	13.7	10.8	10.9	<b>9.8</b>	17.9	17.1	17.3
Команды Яндекса	4.4	4.3	3.8	19.5	12.2	18.6	21.8	11.2	<b>1.9</b>	3.4	7.1	12.2	8.7
Звонки заказы	20.0	18.8	22.8	32.5	35.7	23.7	24.8	21.8	15.5	<b>13.7</b>	27.9	18.5	29.5
Звонки поддержка	12.9	12.6	23.8	29.4	34.3	26.8	27.5	23.7	14.2	<b>12.4</b>	16.8	14.8	28.9
Среднее	11.02	10.69	13.95	16.02	20.24	16.21	16.84	13.78	8.64	<b>8.42</b>	14.67	12.79	15.99

## Применение регрессионного подхода.

Регрессия позволяет моделировать связь между объективными характеристиками речи и субъективными оценками слушателей. Примером работ в этой области является [Guéguin M. et al. A step further to objective modeling of conversational speech quality //2006 14th European Signal Processing Conference. – IEEE, 2006. – С. 1-5.]. В работе применена множественная линейная регрессия, предсказания оценки качества речи MOS на основе параметров речевого сигнала.





# Комплексирование полученных методов оценивания (

Этап направлен на объединение результатов работы по отдельным факторам в единую оценку. Комплексирования необходимо привести отдельные метрики к единому масштабу метрик качества в диапазону [0;1], при этом значения, выходящие за границы данного диапазона в процессе эксплуатации, могут заменяться на соответствующие граничные значения.

Математическое объединение выходов отдельных факторов, с возможным применением коэффициентов.

- 1) взвешенное среднеарифметическое;
- 2) среднее усечённое;
- 3) невзвешенное среднеарифметическое;
- 4) принятие итогового решения с помощью методов машинного обучения из библиотеки scikit-learn (SVM, CART, Random forest, Ada Boost, NB, QDA, MLP).

Для определения весов могут использоваться подходы, связанные с оцениванием информативности отдельных компонент и их влияния на принятие итогового решения, такие как методы анализа дающих для определению важности определённой характеристики для принятия итогового решения при использовании конкретных классификаторов.





# Комплексирование полученных методов оценивания (

При наличии индивидуальных рекомендаций по значимости отдельных критериев качества. На приоритет разборчивости, как определяющей возможность речевого взаимодействия, сравнительная низкая значимость естественности и узнаваемости. Данные значения весов могут быть получены эксперты путем, что не влияет на возможность получения итоговой оценки качества речи.



**Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»**  
**25-28 ноября 2025, Ставрополь, Россия**



Основные этапы метода комплексирования:

1. Этап предобработки: преобразование речевых сигналов к формату, пригодному для обработки с использованием рассматриваемых методов оценивания отдельных аспектов качества речи, и представленному в рамках используемого набора данных.
2. Этап выбора методов оценивания отдельных аспектов качества речи: разборчивости, узнаваемости и естественности.
3. Этап применения выбранных методов: получение оценок отдельных аспектов качества речи: разборчивости, узнаваемости, естественности, при необходимости - их нормирование к диапазону [0; 1] для дальнейшей обработки.
4. Этап выбора подхода к комплексированию оценок: на основе вариантов комплексирования, их положительных и отрицательных особенностей.
5. Этап выбора важности отдельных аспектов качества речи: при наличии в критерии весовых коэффициентов отдельных аспектов качества речи используется метод анализа иерархий для получения оценок этих коэффициентов.
6. Этап комплексирования: получение итоговой единой комплексной интегральной оценки качества речи на основе выбранного метода комплексирования и его параметров, при необходимости - нормирование итоговой комплексной оценки.



Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия



## Варианты комплексирования

Название метода	Преимущества	Недостатки
Линейная комбинация	Простота реализации и интерпретации	Требует точного подбора весовых коэффициентов
Мультипликативная модель	Учитывает синергический эффект критериев	Может быть чувствительным к низким значениям хотя бы одного из критериев
Среднее геометрическое	Учитывает взаимозависимость критериев и снижает влияние выбросов	Менее интуитивно понятен чем линейная комбинация
Среднее гармоническое	Акцентирует внимание на низких значениях критериев	Может быть излишне строгим
Минимально-максимальное комплексирование факторов	Подчёркивает важность баланса между критериями или акцентирует внимание на сильных сторонах	Игнорирует вклад других, если один из них низкий или может игнорировать слабые стороны
Среднее логарифмическое комплексирование	Учитывает нелинейные зависимости	Менее интуитивно понятен
Среднее экспоненциальное комплексирование	Подчёркивает высокие значения критериев	Может быть чувствителен к выбросам

Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия





## Варианты комплексирования

Название метода	Преимущества	Недостатки
Произведение с весами	Гибкость в учёте важности критериев	Сложность подбора весов
Машинное обучение	Учёт сложных нелинейных зависимостей	Требует большого объёма данных
Функция полезности	Максимальная гибкость	Требует сложной настройки и понимания предпочтений
Усреднение через лепестковую диаграмму	Простая геометрическая интерпретация с точки зрения лепестковой диаграммы	Неочевидное поведение при отсутствии нормировки
Медоид	Устойчив к выбросам, получаемое значение принадлежит выборке	Отсутствие статистической эффективности



Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия



Пример. Есть набор оценок: разборчивость – 0,811, естественность – 0,644, узнаваемость – 0,715. Для свертки взвешенное гармоническое среднее, т.к. позволяет гибко учитывать важность критериев и позволяет подчеркнуть важность каждого из критериев в отдельности (отдельное значение компоненты, равное 0 обращает весь критерий в 0). Экспертная оценка: разборчивость важнее узнаваемости в 3 раза. Разборчивость важнее естественности в 5 раз. Узнаваемость важнее естественности в 2 раза.

Матрица попарных сравнений:

Разборчивость Узнаваемость Естественность

Разборчивость	1	3	5
Узнаваемость	1/3	1	2
Естественность	1/5	1/2	1

Нормализация матрицы

Столбец 1:  $1+1/3+1/5=1.533$

Столбец 2:  $3+1+0.5=4.5$

Столбец 3:  $5+2+1=8$

Делим каждый элемент на сумму столбца:

Разборчивость Узнаваемость Естественность

Разборчивость	0,652	0,667	0,625
Узнаваемость	0,217	0,222	0,25
Естественность	0,130	0,111	0,125

Оценки важности как средние по строкам: Разборчивость – 0,648 Узнаваемость – 0,230, Естественность – 0,122.

Есть набор оценок: разборчивость – 0,811, естественность – 0,644, узнаваемость – 0,715.

Матрица попарных сравнений:

	Разборчивость	Узнаваемость	Естественность
Разборчивость	1	3	5
Узнаваемость	1/3	1	2
Естественность	1/5	1/2	1

Оценки важности как средние по строкам: Разборчивость – 0,648 Узнаваемость – 0,230, Естественность – 0,122.

Для оценки согласованности: умножаем исходную матрицу на веса и делим на веса:

$$1 \cdot 0.648 + 3 \cdot 0.230 + 5 \cdot 0.122 \approx 1.994$$

$$(1/3) \cdot 0.648 + 1 \cdot 0.230 + 2 \cdot 0.122 \approx 0.693$$

$$(1/5) \cdot 0.648 + (1/2) \cdot 0.230 + 1 \cdot 0.122 \approx 0.368$$

$$\lambda_{\max} = 0.648/1.994 + 0.230/0.693 + 0.122/0.368 \approx 3.08$$

$$\text{Индекс согласованности (ИС) ИС} = (\lambda_{\max} - n)/(n - 1) = (3.08 - 3)/2 = 0.04$$

Отношение согласованности (ОС) для матрицы 3×3 случайная согласованность (СС) = 0.58.

ОС=ИС/СС=0.04/0.58 ≈ 0.069(6,9%) Вывод: Поскольку ОС < 10%, суждения эксперта согласованы.

Итоговая оценка качества  $Q=1/(0,648/0,881+0,230/0,644+0,122/0,715)=0,792$ .

# Эксплуатация построенного метода (6)

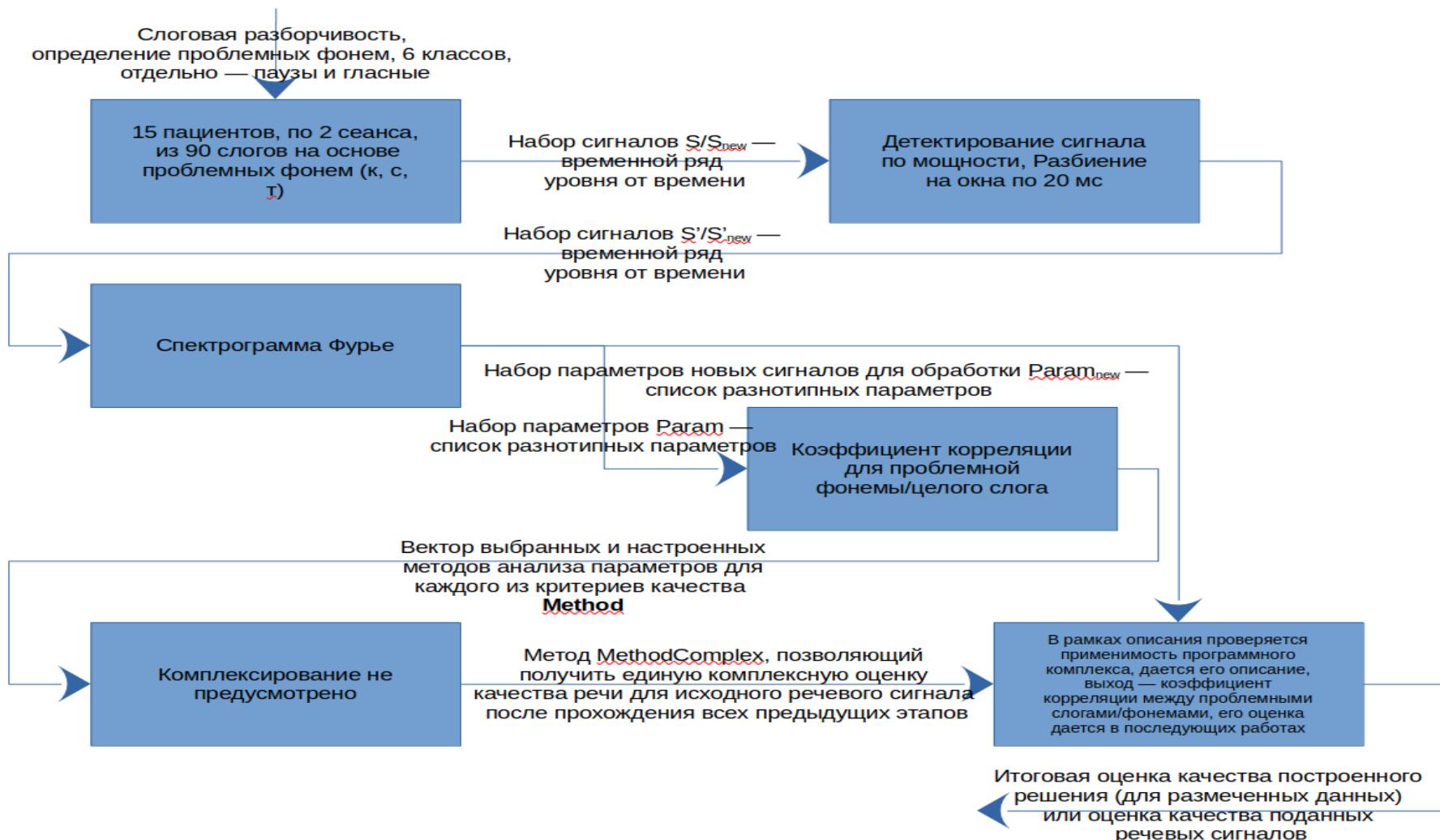
Настройка параметров запуска оценивания, расчет метрик качества для новых значений, валы выбранных методов анализа и моделей, построенных на их основе, сравнение результатами, полученными на данных, используемых при построении, с результатами, полученными другими авторами для различных отдельных аспектов качества и их комбинаций. Анализ полученных результатов, экспериментальное уточнение, рекурсивная модификация предыдущих этапов.



Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»  
25-28 ноября 2025, Ставрополь, Россия

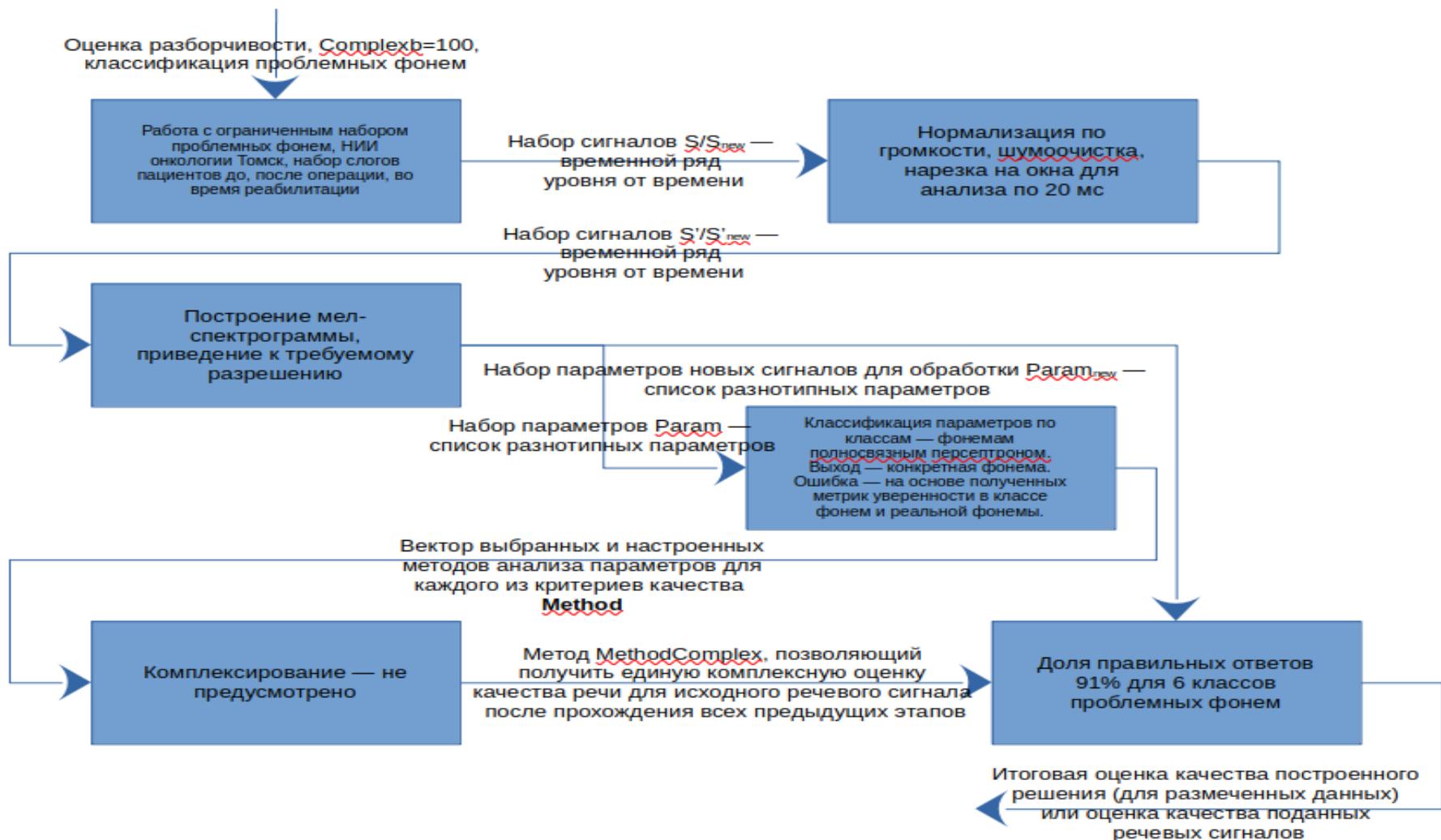
# Слоги. Расстояния.

4. Novokhrestova D., Kostyuchenko E., Kosenko E. Automation of speech quality assessment in speech rehabilitation // International Journal of Advanced Trends in Computer Science and Engineering. – 2019. – T. 8. – №. 6. – С. 3026-3030.



## Слоги. Классификация по фонемам.

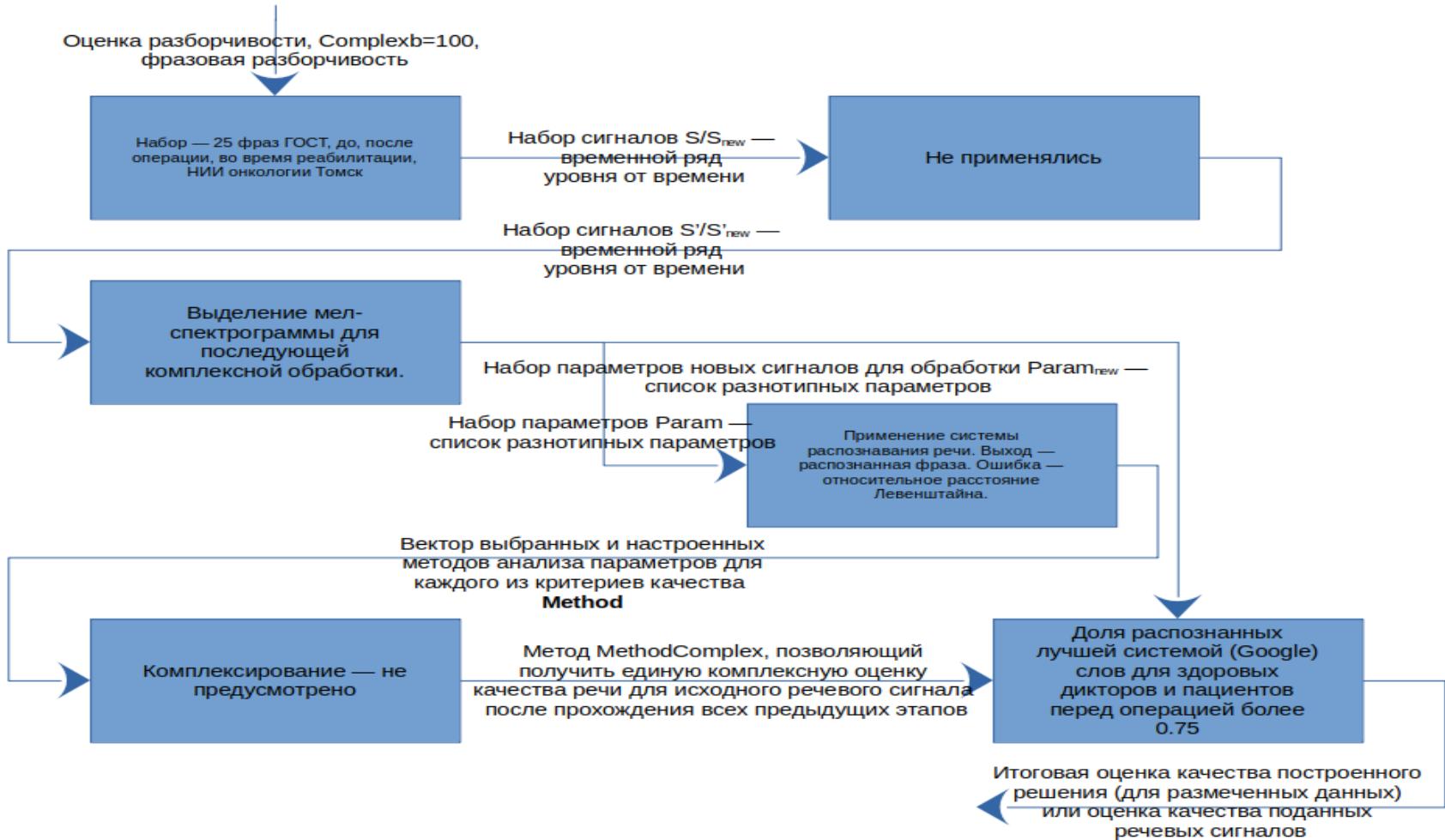
14. Kostuchenko E, et al. Assessment of syllable intelligibility based on convolutional neural networks for speech rehabilitation after speech organs surgical interventions //International Conference on Speech and Computer. – Cham : Springer International Publishing, 2019. – С. 359-369.





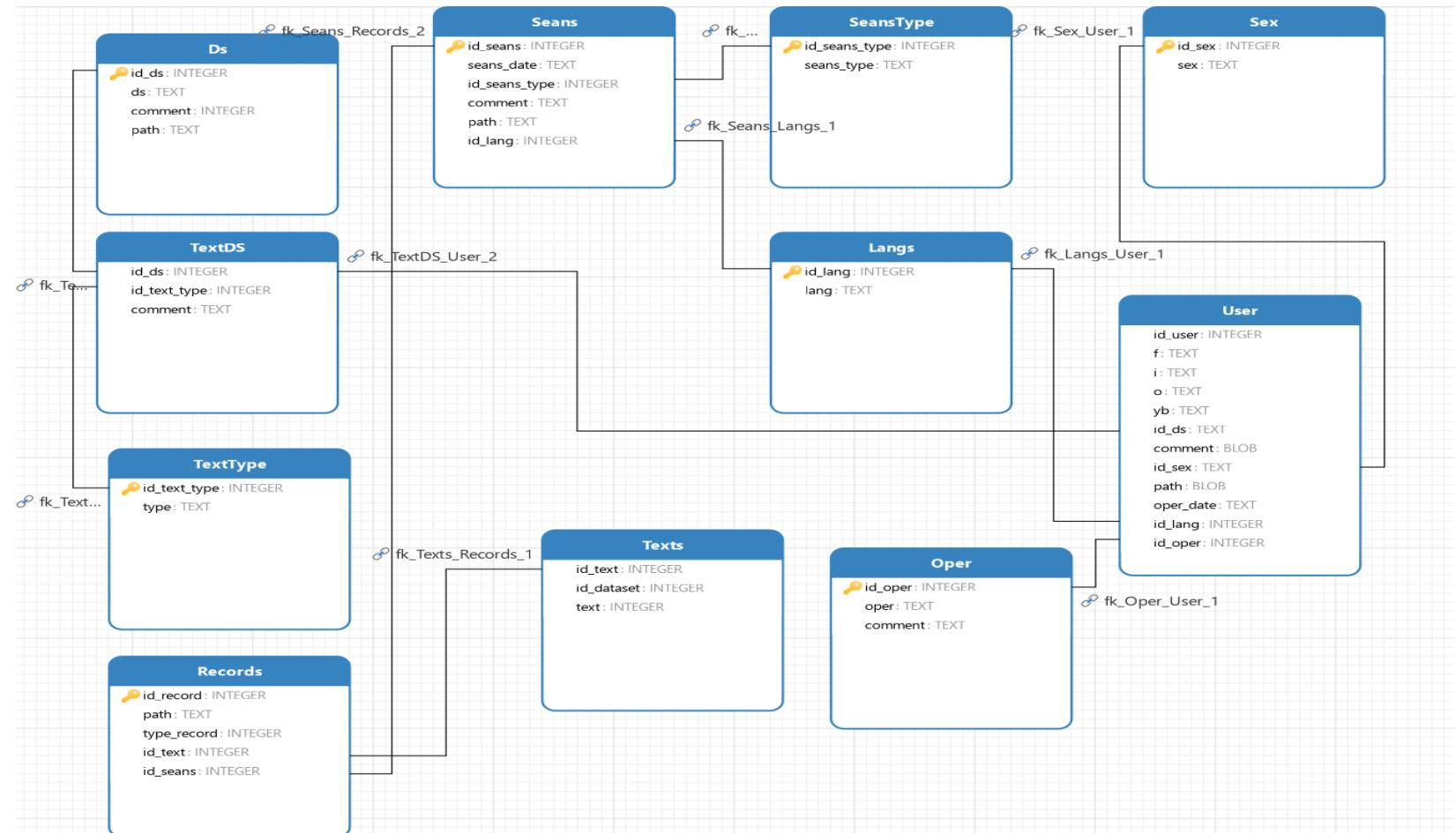
## Слоги. Распознавание фонем через классификацию.

15. The evaluation process automation of phrase and word intelligibility using speech recognition systems [Текст] / E. Kostuchenko [et al.] // International Conference on Speech and Computer. — Springer. 2019. — P. 237—246.





Общий объем набора – более 120 дикторов, 4 категории оперативных вмешательств сеансов реабилитации, записи для оценивания фразовой и слоговой разборчивости, более записей.





## Система звукозаписи и оценивания

Разборчивость речи

N	Фамилия Имя Отчество	Год рождения	Диагноз	Лечение	Логопедически диагноз
13	[REDACTED]	1961	ФФИК шеи ...	Рак языка Т...	Послеоперац...
14	[REDACTED]	1950	13.06.2013 У...	Злокачеств...	
15	[REDACTED]	1956	ФФИК шеи ...	Рак языка Т...	Послеоперац...
16	[REDACTED]	1959	ФФИК шеи ...	Рак языка с...	Дизартрия
17	[REDACTED]	1954	14.11.2013Ф...	Рак языка с...	Послеоперац...
18	[REDACTED]	1959	22.04.2014 Т...	Рак ротовоглот...	Послеоперац...
19	[REDACTED]	1949	ФФИК шеи с...	Рак языка N...	Дизартрия
20	[REDACTED]	1947		Рак слизист...	
21	[REDACTED]	1946	27.08.2014Оп...	Рак языка Т...	
22	[REDACTED]	1977	ФФИК шеи ...	Рак языка Т...	Послеоперац...

К сессиям

Сессии

N сессии	Тип записей	N пациента	Дата	Средняя оценка
786	1	44	31.08.2015	
787	2	44	31.08.2015	
797	1	44	03.11.2015	
798	2	44	03.11.2015	
799	1	44	13.11.2015	
800	2	44	13.11.2015	

Создать новый сессии

Перейти к сессии

Запись слогов

Пациент	Сергеев Сергей Сергеевич	Количество слогов	100
id сессии	950	Запись	
Тип записи	Слоги ГОСТ		
Дата	14.09.2017		
		<b>Начать запись</b>	
		<b>Следующий слог</b>	
		<b>Завершить запись сессии</b>	

Слоговая разборчивость

N записи	N сессии	N слога	Расположение	Слышимый слог
2277	138	15	d:\sound\Slog	[REDACTED]
2278	138	16	d:\sound\Slog	[REDACTED]
2279	138	17	d:\sound\Slog	[REDACTED]
2280	138	18	d:\sound\Slog	[REDACTED]
2281	138	19	d:\sound\Slog	[REDACTED]
2282	138	20	d:\sound\Slog	[REDACTED]
2283	138	21	d:\sound\Slog	[REDACTED]
2284	138	22	d:\sound\Slog	[REDACTED]

Запись слогов

Пациент	[REDACTED]	Количество слогов	100
id сессии	951	Запись	
Тип записи	Слоги ГОСТ		
Дата	14.09.2017		
		<b>Начать запись</b>	
		<b>Следующий слог</b>	
		<b>Завершить запись сессии</b>	

НЯК

н'ак



## Заключение

1. Проведен анализ и классификация существующих подходов к оцениванию качества речи, выявлены несуществующие подходы.
2. Построена методология, направленной оценивание качества речи в процессе проведения речевой реабилитации.
3. Сформирован, структурирован и собран набор данных для проведения исследования в области оценивания различных аспектов качества речи.
4. Проведено исследование аспектов, связанных с предобработкой сигналов и существующими методами оценивания различий между сигналами для оценивания качества речи (через расстояние, классификацию и распознавание).
5. Проведено применение методологии для оценивания качества речи через разборчивость, узнаваемость и естественность на примере решаемых исследовательских задач оценивания качества речи, а также через разработанный метод их комплексирования.
6. Разработан единый программный комплекс для оценивания качества речи.



**Международная школа молодых ученых «Искусственный интеллект и кибербезопасность»**  
**25-28 ноября 2025, Ставрополь, Россия**



**Спасибо за внимание!**

**Evgeny Kostyuchenko**

Ph.D., Associate Professor

Deputy Head of the Laboratory of Biomedical Research (TUSUR University)

40, Lenina st., Tomsk, Russia, 634050

ph.: +7 (3822) 413426

E-mail: [key@keva.tusur.ru](mailto:key@keva.tusur.ru)

Web: <http://www.tusur.ru>